LCA Methodology

GM-Troph

A Low Data Demand Ecotoxicity Effect Indicator for Use in LCIA

Henrik Fred Larsen* and **Michael Hauschild**

Department of Manufacturing, Engineering and Management, Technical University of Denmark (DTU) Building 424, 2800, Lyngby, Denmark

* **Corresponding author** (hfl@ipl.dtu.dk)

Abstract

Goal, Scope and Background. The development of ecotoxicity effect indicators (EEIs) for use in life cycle impact assessment (LCIA) has only been going on for about two decades. Traditionally, no-effect indicators have been applied. In this paper we focus on the development of an effect-based (i.e. EC_{50}-based) average indicator, the GM-troph. The indicator is estimated by use of the hazardous concentration for 50% of the covered species ($HC50_{EC50}$) and is designed to work on a low substance data availability of only three acute data values, which is often required in LCIA.

Methods. The study includes a theoretical description and a test on real data of three different effect-based average approaches (arithmetic mean, geometric mean and median) focusing on their statistical robustness. The data set used for the testing is composed of real ecotoxicity effect data for eleven different substances representing seven different toxic modes of action (TMoA).

Results and Discussion. The theoretical considerations and the test on real data show that the geometric mean is the most robust average estimator for $HC50_{EC50}$, especially in the frequent situation where data availability is limited to a few data points. Test results indicate that in some cases of unequal representation of the different taxa (or trophic levels) in the underlying data set, estimations of average toxicity (i.e. $HC50_{EC50}$) may be biased if each single test data (at a species level) is used as data points instead of averages at trophic levels.

Conclusions and Recommendations. Based on these results, the following recommendations are given for the choice of estimation principle for the EEI: The indicator shall be based on $HC50_{EC50}$ estimated as the geometric mean of three (average) EC_{50} values, covering the three main taxa, plants, invertebrates and vertebrates, which represent the three trophic levels of the ecosystem, primary producers, primary consumers and secondary consumers. In practice, the EEI shall be based on data from laboratory tests with algae, invertebrates (crustaceans) and fish. Instead of using the often wide 95% confidence limits, it is recommended to use the range given by the observed maximum and minimum values as limits around the $HC50_{EC50}$. Further, it is recommended to use EC_{50}(chronic) values when possible. Often, only acute data will be available, and here the use of best estimate assessment factors is recommended to extrapolate from acute to chronic values. As a starting point, an acute to chronic ratio of 2 between $HC50_{EC50}$(acute) and $HC50_{EC50}$(chronic) is recommended, but more research is certainly needed in this area. Due to the comparative framework of LCIA it is recommended only to use test results from laboratory tests, fulfilling certain standard conditions, i.e. applying standard organisms, and measuring well defined endpoints over restricted test durations.

Perspectives. The ability of a geometric, mean-based $HC50_{EC50}$ to represent the toxicity of very toxic substances and toxicity towards very sensitive species has not been dealt with here, and further research is needed. However, on the basis of the results from the tests on real data, it may be anticipated that the GM-troph with its max-min limits to some degree accounts for the toxicity even to the most sensitive species among standard organisms, if representative toxicity data are available.

Glossary

AF	Assessment Factor
EC_{50}	Effect Concentration (50% of test organism affected)
EDIP	Environmental Design of Industrial Products
EEI	Ecotoxicity Effect Indicator
GM	Geometric Mean
HC50	Hazardous Concentration for 50% of included species
LC_{50}	Lethal concentration (50% of test organism dead)
LCA	Life Cycle Assessment
LCIA	Life Cycle Impact Assessment
LOEC	Lowest Observed Effect Concentration
$LowEC_{50}$	Lowest EC_{50}
NEC	No Effect Concentration
NOEC	No Observed Effect Concentration
OECD	Organisation for Economic Co-operation and Development
OMNIITOX	Operational Models aNd Information tools for Industrial applications of eco/TOXicological impact assessments
PAF	Potentially Affected Fraction of species
PNEC	Predicted No Effect Concentration
PVF	Potentially Vanished Fraction of species
QSAR	Quantitative Structure Activity Relationship
SSD	Species Sensitivity Distribution
TGD	Technical Guidance Document
TMoA	Toxic Mode of Action

Introduction

This paper is based on the report 'Implementation of the OMNIITOX Base Model. Part VIII – Implementation of the ecotoxicological effects module' by Larsen et al. (2004), which was done as part of the OMNIITOX[1] project. The indicator described here, the GM-troph, is used in the OMNIITOX base model for calculating ecotoxicity characterisation factors.

[1] **OMNIITOX** is an EU-project under the Competitive and Sustainable Growth Programme, running from 2001 to 2005. OMNIITOX will facilitate decision-making regarding potentially hazardous compounds by improving methods and developing information tools necessary for impact assessment of toxic chemicals within LCA and risk assessment. Project partners are Technical University of Denmark; Leiden University, The Netherlands; University of Stuttgart, Germany; École Polytechnique Fédérale de Lausanne (EPFL), Switzerland; Chalmers University of Technology, Sweden; European Chemicals Bureau, JRC, Ispra, Italy; Volvo Technology Corporation, Sweden; Procter & Gamble EUROCOR, Belgium; Stora Enso AB, Sweden; Antonio Puig, S.A. Spain; Randa Group S.A, Spain. More information about OMNIITOX can be found at http://omniitox.imi.chalmers.se/OfficialMirror.

In recent years, research on ecotoxicity effect indicators (EEIs) has intensified. The general frame has been defined by the SETAC Task Group on Ecotoxicity (Hauschild and Penning–ton 2002, Pennington et al. 2004). Further in the development of EEIs, different ways have been explored for the estimation of average toxicity and its associated uncertainty, using non-parametric statistics (Payet and Jolliet 2005) or using parametric statistics (Payet 2004, 2005). Most recently, an evaluation of different EEI approaches performed by Larsen and Hauschild (2006) puts focus on data availability. This paper describes the development of an unbiased, average estimate of the toxicity of chemicals based on a modest data set, and suggests an improvement of the selection of the input data, principles of estimation, and calculation of the uncertainty for the effect indicator.

Given the general framework of life cycle assessment (LCA) and especially life cycle impact assessment (LCIA) (ISO 2000; Udo de Haes and Lindeijer 2002), and previous work done on EEIs, a number of general constraints with relevance for the EEI approaches may be outlined:

- A general condition for the LCIA models is that the impact category indicator must be additive.
- In contrast to (tiered) risk assessment the indicator shall be a best estimate, i.e. not a conservative estimate.
- Emission of a toxicant mapped in a life-cycle inventory (LCI) is regarded as a single pulse without time duration and, therefore, time and space are integrated in the assessment giving further restrictions to the modelling.
- In ordinary LCAs the location of the processes which release toxicants to the environment is usually not precisely known and, therefore, site-specific models cannot easily be used. Most often we have to rely on large-scale averages of environmental conditions.
- The large number of substances covered by an LCI calls for a model that relies on relatively few input data in order to make the data gathering feasible.
- The availability of ecotoxicological effect data for the majority of chemicals on the market puts severe restrictions on the data demand of the EEI model.
- The assessment of impact (or damages) on ecosystems in LCA requires compatibility between the fate model and the effect model. The two models can, for example, be connected as shown in **Eq. 1** for the characterisation factor (CF) per kg emitted contaminant, indicating the fraction of species experiencing an increase in stress for a change in contaminant concentration. The 'effect part' (i.e. the effect indicator, 0.5/HC50) is here expressed in PAF·m³·kg⁻¹ and the 'fate part' (i.e. change in concentration, dC) can then be expressed in kg m⁻³ per kg emitted contaminant. In this case, the characterisation factor is expressed in PAF per kg emitted contaminant. If the 'fate part' is expressed in a time integrated fraction of the emitted amount (days), as in the OMNIITOX base model, the characterisation factor is expressed in PAF m³ days per kg emitted contaminant.

Two main directions can be identified for the indicator, viz. the predicted no-effect concentration (PNEC) approach and the potentially affected fraction of species (PAF) approach (Larsen and Hauschild 2006). Both have their pros and cons.

The PNEC approach, as used today, has a low data demand with high data availability, but results in a conservative estimate aiming at protecting the most sensitive species in the ecosystem. The PAF approach, as used today, typically results in non-conservative estimates, but has a relatively high demand on data with low availability (see Larsen and Hauschild (2006) for detailed discussion on PNEC and PAF). In the comparative context of LCA, where best estimates are sought, the choice of a PAF approach based on average toxicity seems preferable, and further secures the (at least theoretical) possibility to interpret the results in terms of damage. This is important within LCA because it makes it possible to gather different classes of midpoint impact categories in one common endpoint category, for example the impact to ecosystems from land-use (expressed in units of biodiversity) and the ecotoxic impact to ecosystems (expressed in units of PAF). A damage model may translate the midpoint indicators into an endpoint or damage indicator expressing them in equivalent units of potentially vanished fraction of species (PVF). The compatibility between the average toxicity assessment approaches described in this paper and the existing damage models has been studied and possibilities identified to translate the midpoint indicator based on the geometric mean into a damage indicator expressed as change in biodiversity (i.e. PVF) (Payet and Larsen 2006). This is a key issue for further methodological development in LCA, especially for link between the land-use, eutrophication, acidification and the impact due to toxicants (ecotoxicity).

As data availability is low in LCIA (typically three acute data values), an EEI based on average toxicity (i.e. $HC50_{EC50}$) with a working point at PAF = 0.5 on the PAF curve seems most reasonable (Larsen and Hauschild 2006). A characterisation factor (CF) based on this type of indicator can be expressed in the following way (per kg emitted substance):

$$CF = dPAF = EEI \cdot dC = \frac{0.5}{HC50_{EC50}} \cdot dC \qquad (1)$$

The $HC50_{EC50}$ is the hazardous concentration at which 50% of the included species have their EC_{50} value exceeded (endpoint, for example, mortality).

The main advantages and disadvantages of choosing such an effect-based as compared to a no-effect based indicator (i.e. PNEC based) are given below:

Advantages:
- The risk of bias from the laboratory test set-up is low compared to a no-effect based indicator, where the highest tested concentration, which is not statistically different in toxicity from the control concentration, is typically reported.
- The use of a value which is estimated and placed in the centre of the concentration response curve (i.e. HC50) gives the lowest uncertainty.
- The quantification of damage in terms of potential loss of species is possible (at least in theory).

Disadvantages:
- The focus is shifted away from protection of the function and structure of ecosystems.
- The importance of very sensitive species may be neglected.

A potential bias between ecotoxicity and other impact categories, which model lower levels of impact, may be removed by normalisation.

When estimating ecotoxicity effect indicators applying the PAF approach, the data used and the principles applied when determining the HC50 are crucial to the outcome, as shown by Larsen and Hauschild (2006). The main goal of this paper is to propose and document a recommendable way to estimate the HC50 value from the typical data availability of three acute data values.

1 Estimation Principles

Two different estimation principles are used today for estimation of an effect-based HC50 – the median and the geometric mean (Larsen and Hauschild 2006). In our comparison, we will look at these two average estimation principles and further include the arithmetic mean, and the alternative no-effect based PNEC approach (here represented by the lowest EC_{50}, termed $LowEC_{50}$ on which the PNEC would be based) for comparative reasons. First, the different estimation principles are evaluated from a theoretical point of view, and in Section 5, a test on real data is carried out.

The four estimation principles are:

- arithmetic mean (effect-based)
- geometric mean (effect-based)
- median (effect-based)
- PNEC (no-effect based)

The arithmetic mean (termed mean) is certainly the most used estimator of average in general, but should not normally be used for calculating the average effect concentration within ecotoxicology because it presumes that the data set is normally distributed, which is typically not the case for toxicity (see below).

The geometric mean is typically used when estimating average toxicity for a population, e.g. in a standard laboratory test. According to Newman and Dixon (1996), an individual lethal dose (individual effective dose (I.E.D.)) exists for each individual within a population and often displays a lognormal distribution because 'in biological material the variation often shows a geometrical rather than an arithmetic distribution' (cited ref. Bliss 1935). If the average within genera (term used in classification meaning a group of closely related species) is used as data input when dealing with species' sensitivity distributions (SSDs), the geometric mean is also typically used (e.g. Versteeg et al. 1999). This is also the case when the HC50 is estimated from the typical SSD approach (e.g. Aldenberg et al. 2002), which is based on one data point (e.g. NOEC or EC_{50}) from each single species in the group of species selected.

The geometric mean presumes lognormal distribution of the data set. It is possible to test for this precondition by use of tests for goodness of fit, e.g. Anderson-Darling test (Aldenberg et al. 2002). However, this is not always done (Forbes and Calow 2002a, Newman et al. 2002), which may be a problem because lack of fit occurs quite frequently as demonstrated by Newman et al. (2002), thus showing that 27 of 51 data sets failed the test for lognormality. But, for the typical case where three pieces of data are available, it will not be possible to perform any goodness of fit test with any confidence.

The failure of the data set for many substances to meet the precondition of lognormal distribution is one of the reasons why Payet and Jolliet (2005) have developed a median-based ecotoxicity effect indicator (AMI) based on non-parametric statistics (median and confidence limits by bootstrapping). However, as stated in Aldenberg et al. (2002) and Newman et al. (2002), the use of non-parametric statistics (i.e. median based) requires a larger data set than if parametric statistics (i.e. a geometric mean) is used. Anyway, according to the method by Payet and Jolliet (2005), the calculation of confidence limits around the $HC50_{EC50}$, estimated by bootstrapping, demands at least 5 data points and if only three are available (as the typical case here), the difference between the median and the two extremes is multiplied by a factor of two (assumed factor from extrapolation) and used as confidence limits (Payet et al. 2002). As an alternative, a parametric version of AMI making use of the geometric mean has been developed (Payet 2004, 2005) and implemented in the LCIA method IMPACT 2002+ (Jolliet et al. 2003).

Even though the PNEC approach is used as the dominating approach within generic risk assessment today (e.g. EC 2003a) and still very much used in LCIA (e.g. Huijbregts et al. 2000, Hauschild et al.1998), it is not directly included here because it is not an average approach. Only the lowest EC_{50} value ($LowEC_{50}$), on which the PNEC is typically based, by use of an assessment factor, is included for comparative reasons.

1.1 Theoretical examples

In the following theoretical examples, it is anticipated that the test organisms included in the data set for a substance cover three trophic levels, represented by algae, crustaceans and fish (see Fig. 4 and Section 3 for definition of trophic levels, etc.).

If we look at the two theoretical cases (chemical 1 and chemical 2) in **Fig. 1**, it is obvious that the median is not able to distinguish between the average toxicity of the two chemicals even though chemical 1 is very toxic to one of the tested
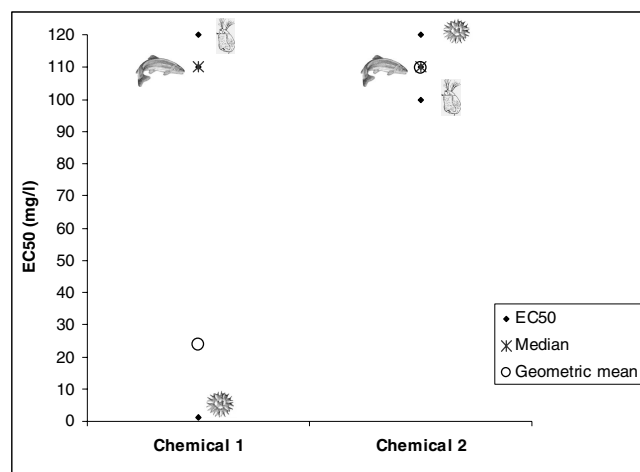

Fig. 1: Two theoretical cases illustrating the difference between the ability of the median and the geometric mean to reflect average toxicity of three EC_{50}s
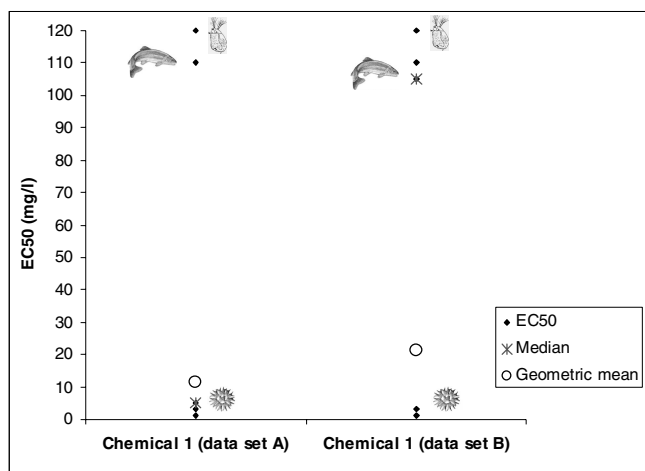
Fig. 2: Two theoretical cases of different data availability for chemical 1 illustrating the ability of the median and the geometric mean to reflect average toxicity of five EC_{50}s

Fig. 3: Two theoretical cases illustrating the inability of $LowEC_{50}$ to distinguish between chemicals with different toxicity distributions

organism (algae), whereas chemical 2 demonstrates a low toxicity to all three tested organisms. The geometric mean, however, is able to distinguish between the average toxicity of the two chemicals.

Looking at chemical 1 only and now assuming two cases of different data availability as shown in **Fig. 2**, the median shows a low robustness. If the fifth data point represents one of the two not very sensitive organism groups instead of the most sensitive organism group (e.g. algae for herbicides), the median jumps from a relatively low EC_{50} value to a relatively high EC_{50} value.

Even with the relative robustness of the geometric mean illustrated here (see Fig. 2), there may still be a problem in using toxicity data for individual species rather than a geometric mean for each of the organism groups or taxa (general term for taxonomic units in classification), which are taken as representatives for each of the trophic levels. As pointed out by Aldenberg et al. (2002), a biased species selection will lead to biased, estimated parameters (e.g. HC50, estimated as the geometric mean) in SSDs. In some cases we may thus have many test data for one trophic level, and only one (or a few) data for the two other trophic levels. Just calculating the geometric mean of these data (i.e. at the species level) may lead to bias putting a weight on the trophic level with many measured values, which may be disproportionate to the ecological relevance of this trophic level and instead reflect the fact that, through regulation, it has been given a high priority in ecotoxicity testing. Furthermore, 'it is generally assumed that members of the same taxonomic group are more similar to each other in sensitivity than to members of other taxonomic groups' (Forbes and Calow 2002a). Nevertheless, studies on SSDs are typically based on a haphazard collection of species with doubtful ecological relevance as shown by Forbes and Calow (2002a).

In **Fig. 3**, a theoretical example of two chemicals (chemical 3 and chemical 4) with quite different toxicity profiles illustrates the inability of distinguishing by the PNEC approach (here, lowEC$_{50}$). Note that PNEC is typically a factor 100–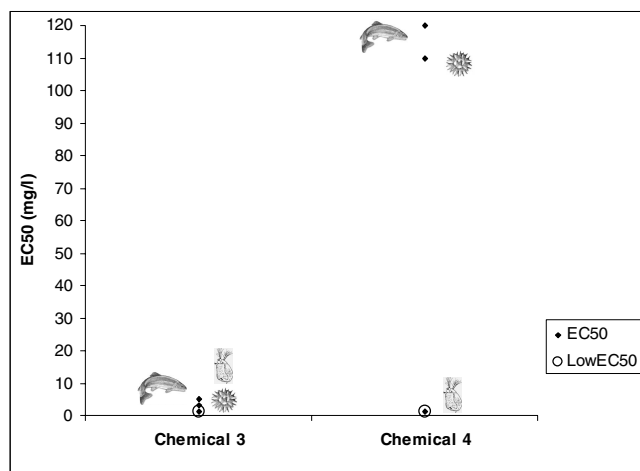1000 lower than LowEC$_{50}$ depending on the assessment factor used (Hauschild et al. 1998, EC 2003a). As can be seen in Fig. 3, the $LowEC_{50}$ is the same for chemical 3 and chemical 4. However, chemical 3 is very toxic to all three organisms tested, whereas chemical 4 is only very toxic to one of the organisms tested (i.e. the Daphnia). The PNEC approach is aiming at protecting the most sensitive species (hereby inherently assuming that the ecosystem health is protected) and is therefore not an average approach. Anyway, it seems reasonable to anticipate that the toxic pressure (and hereby the effect) on an ecosystem will be higher if the chemical is very toxic to organisms from all three trophic levels (chemical 3) than if it is only toxic to one trophic level or one species (chemical 4). It is possible that this one sensitive species is a key species and, therefore, very important for the structure and function of the ecosystem, but this will probably only be true in very few cases if at all, and is not the average situation which we aim for in LCIA in order to avoid a consistent bias caused by basing the ecotoxicity indicator on conservative assumptions.

2 Environmental Compartments

In this paper, we will focus on the freshwater pelagic compartment for which the data availability is best. A proposal on how to deal with other compartments can be found in Appendix A (see OnlineEdition).

3 Trophic Levels

An ecosystem can be characterised by its structure and function. Structural descriptors like species diversity (species richness) and trophic structure, and functional descriptors like biomass production, energy flow and nutrient recycling, may be used (Cairns et al. 1995, Sand-Jensen 2000, Pratt and Cairns 1996). Here, we focus on the trophic structure because it includes both the different life forms (e.g. autotrophic and heterotrophic organisms) and, to a high degree, different feeding types (e.g. herbivorous and carnivorous), and to some degree the level of biological evolutionary development/complexity (e.g. micro-organism, invertebrates and vertebrates) in a structured way. Furthermore, the legal
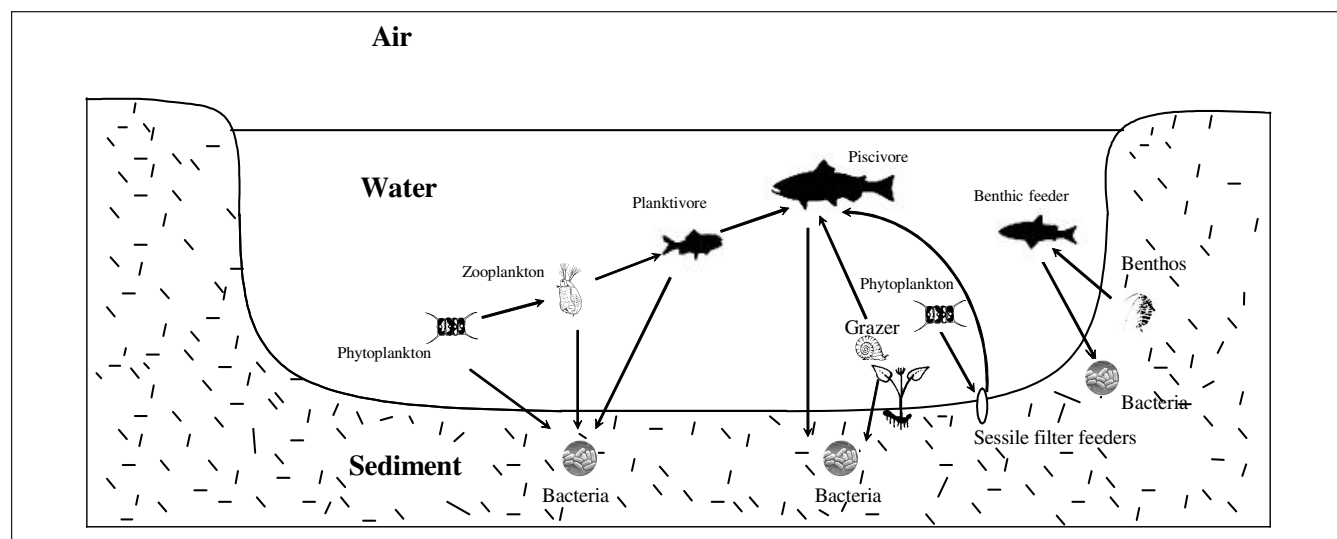
Fig. 4: Simplified food web for a freshwater ecosystem. Adapted/modified from Chapman et al. (2003), with permission

framework for the dominating part of the available eco-toxicity test data is based on the representation of trophic levels (e.g. EC 2003a).

A simplified food web with food chains for a freshwater ecosystem is shown in **Fig. 4**. The positions along the food chains are trophic levels (Whittaker 1975). In the freshwater ecosystem the different trophic levels are represented, for example, by the phytoplankton (primary producers), the zooplankton (primary consumers, herbivores), the planktivore fish (secondary consumers, first carnivores) and the piscivore fish (tertiary consumers, secondary carnivores).

As is evident from the description in Appendix B (see OnlineEdition), the laboratory ecotoxicity effect data for freshwater is dominated by test on these 3–4 trophic levels, represented by algae (phytoplankton), crustaceans (zooplankton) and fish. However, as illustrated in Fig. 4, one more major trophic level exists, the reducers, here illustrated by bacteria.

The reducers are represented by micro-organisms, i.e. bacteria and fungi. Even though these organisms are at a low trophic level and may therefore serve as an early warning indicator of potential toxic effects, experience from bioassays (laboratory ecotoxicity tests) seems to indicate a generally lower sensibility of tests on these organisms as compared to tests on organisms representing other trophic levels. In a review on bioassays for microbial systems (Mayfield 1998), several tests measuring different parameters (e.g. growth rate, biomass, and enzyme activity) are described. Most of the evaluated tests are either characterised as not being sensitive enough to reveal responses to toxicity in the natural environment (e.g. ATP assays and microcalorimetric assays) or not developed to a 'standardised' level (e.g. turbidostat culture system). Furthermore, toxicity data are to a great extend lacking except for special chemicals (fungicides for fungi tests). It is therefore not recommended to include the reducers in the calculation of the ecotoxicity effect indicator.

Given these arguments, and considering, as described for instance in the TGD (EC 2003a), that three trophic levels

are typically required within generic risk assessment, thus resulting in higher data availability for these, the following trophic levels are recommended for inclusion in the estimation of HC50:

1. Primary producers (i.e. algae)
2. Primary consumers (i.e. invertebrates)
3. Secondary consumers (i.e. fish)

4 Endpoints

In LCA studies, we typically deal with marginal concentration increases in the 'model' environment, at least if we calculate emissions for a 'small' functional unit. The emissions will therefore mostly give rise to potentially chronic effects, but local emissions leading to potential acute effects may have significant importance as illustrated, for example, by Larsen et al. (2006).

By choosing to work on the level of chronic effects, we run into the problem with a lack of chronic data for the main part of the tested chemicals (Larsen and Hauschild 2006). This problem is normally solved by using acute data to extrapolate to the level of chronic effects by use of assessment factors (AFs). Especially within risk assessment, but also in existing LCIA methods like USES-LCA (Huijbregts et al. 2000) and EDIP (Hauschild et al. 1998), the AFs used are generally considered as conservative and used for estimating the Predicted No Effect Concentration (PNEC). However, a study by Forbes and Calow (2002b) on a limited database indicates that this is often not the case, and that actually a case could be made for increasing the AF used within tiered risk assessment by an order of magnitude, rather than decreasing it. In the same study, it is shown that acute-to-chronic ratio (ACR), on which the AFs are based, can vary strongly between substances. An ACR (calculated between acute LC_{50} values and chronic LOEC values with growth rate as endpoint) of 100–1000 is thus not unusual (Forbes and Calow 2002b), particularly for substances with a specific toxic mode of action (TMoA) like pesticides. In contrast, a factor of well below or around 10 seems appro-

priate for substances with a non-specific TMoA (narcotics). As we are aiming for best estimates in LCA, we should avoid the use of conservative (and in some cases too low) AFs and instead use best estimate factors. For best estimate extrapolation, we thus need different AFs depending on the kind of TMoA of the substance in question and to support such a differentiation, research is needed. Anyway, a study on the ratio between acute $HC50_{EC50}$ and chronic $HC50_{EC50}$ by Payet (2004, 2005) on 92 different chemicals including both narcotics and pesticides shows ACRs of typically 1–10, and, on average, of around 2. The main reason for the relatively low ACRs in this case is that they are based on comparison between average toxicities (i.e. HC50s). The HC50s used by Payet are estimated on the basis of a species level (not averages on trophic level) where individual species contribute with one data value and non-standard tests/test conditions and non-standard test organism are included.

Fortunately, for many of the known, very toxic substances with specific TMoA (e.g. pesticides), chronic data already exist, but we are still faced with the problem for the main part of existing substances which are poorly investigated, and for which the TMoA as well as the chronic toxicity is unknown. Facing the challenge of choosing between an indicator based solely on acute values, and an indicator based on chronic values and making use of assessment factors to a high degree, it seems preferable to choose the chronic indicator when we are aiming at obtaining a best estimate while taking the framework of LCA into account.

As argued earlier, the preferred endpoint for acute test results to be included in calculation of the EEI is an effect concentration, i.e. EC_{50}. EC_{50} (LC_{50}; IC_{50}) is defined as the test concentration of the substance where 50% of the test organisms die (lethal concentration, i.e. LC_{50} for fish, for instance), are immobilized by 50% (immobilized effect concentration, i.e. EC_{50} for *Daphnia*, for example) or growth inhibited by 50% (growth inhibition concentration, i.e. IC_{50} for algae) within predefined time durations in laboratory tests (see Appendix B.2, OnlineEdition, for more details).

For chronic tests, the preferred endpoint is also EC_{50}, but the types of effects measured here are chronic, e.g. inhibition of reproduction and reduced growth. For algae the endpoint (and the value) is the same as for the acute part, i.e. 50% inhibition of growth.

As the no observed effect concentration (NOEC, here defined as chronic NOEC) is the key value for estimating PNEC in risk assessment, the results from very many chronic ecotoxicity studies have been published as NOECs. However, according to the standard test guidelines, also EC_{50} (chronic) should be recorded (at least for *Daphnia*) which is also done and published in some cases (ECOTOX 2003, IUCLID 2000).

Since the number of chronic EC_{50} values which are available is limited compared to the number of NOECs, the estimation of missing EC_{50}(chronic)-values from NOECs is tempting, but also problematic. The NOEC value can be defined as 'the highest concentration of the test substance that produces no statistically significant adverse effect on the exposed population of test organism when compared to an untreated control' (Chapman et al. 1996). It is therefore determined by hypothesis-testing, leading to the fact that the value of NOEC needs to be one of the test concentrations. The determined value for NOEC is therefore dependent on the choice of test concentrations and the experimental variability (i.e. choice of test design), and may be more or less close to the 'true' highest concentration producing no adverse effect or the 'true' no effect concentration (NEC) (Solbé 1998). Several other arguments for the unsuitability of NOEC as a precise, unbiased estimate of toxicity are given in Chapman et al. (1996), including the fact that calculation of confidence intervals is not possible for the NOEC. In contrast, an EC_{50} estimate does not have to be one of the test concentrations, it is not dependent on the precision of the experiment, and a calculation of confidence intervals is possible. Based on these arguments and the fact that we are aiming for an effect-based indicator, it is recommended to avoid the use of NOEC values in the calculation of the ecotoxicity effect indicator.

As stated earlier, the indicator shall as a minimum be able to work on only one acute data value (EC_{50}) from each of the three trophic levels represented by algae, invertebrates (crustaceans) and fish. This minimum situation will be the case for most of the substances with useable data due to generally low data availability as described in Larsen and Hauschild (2006). So, in most cases we are facing the challenge of estimating the most robust $HC50_{EC50}$(acute) based on three acute EC_{50}s and afterwards extrapolate this value to an $HC50_{EC50}$(chronic) value. As just described, best estimate AFs for this extrapolation have unfortunately not been yet developed, and research is needed in this area, which lies outside the scope of this paper. However, as a starting point, a general assessment factor of 2 (Payet 2004, 2005) is recommended.

In the rest of this paper the focus will be on acute ecotoxicity values. General considerations about choice of ecotoxicity tests in the context of LCA, and specific recommendations of test organisms and test criteria to be included are given in Appendix B (see OnlineEdition).

5 Estimation Principles Tested on Real Data

With the aim of finding the most robust effect indicator (i.e. $HC50_{EC50}$) based on only three data values, the different average approaches (i.e. geometric mean, median and mean) are tested on real substance examples below. The substances, which are introduced in **Table 1**, represent different toxic modes of action (TMoA).

Apart from different TMoAs, the substances in Table 1 have been chosen to represent different ranges of acute toxicity within each trophic level or between the trophic levels, and different data availability (as visible from Table 2). As seen from Table 1, the ratio between highest and lowest measured EC_{50} value for a substance varies from a factor of 5 (4-methyl-2-pentanone) to a factor of 20,000 (terbutlyazine).

The TMoA indicated for each of the substances in Table 1 is taken from Russom et al. (1997) and The Pesticide Manual

Table 1: Substances included in the test of different 'average approaches' to estimate $HC50_{EC50}$ with their TMoA and the lowest and highest acute ecotoxicity value found among species of fish, crustaceans and algae

Substance name	CAS No.	Type	TMoA	Lowest EC$_{50}$ (acute) (mg/l)	Highest EC$_{50}$ (acute) (mg/l)
2,3,4,6-Tetrachlorophenol	58-90-2	Intermediate (biocide)	?	0.09	10.1
4-Methyl-2-pentanone	108-10-1	Organic solvent	Base-line narcosis	170	780
2,4-Dichlorophenol	120-83-2	Organic solvent	Base-line narcosis	1.4	21
2-Chloroaniline	95-51-2	Intermediate	Polar narcosis	0.13	150
4-Nitrophenol	100-02-7	Intermediate	Polar narcosis	3.8	32
Dicamba	1918-00-9	Herbicide	Auxin-like growth regulator	0.061	750
Metribuzin	21087-64-9	Herbicide	Photosynthesis inhibitor	0.0117	147
Terbutylazine	5915-41-3	Herbicide	Photosynthesis inhibitor	0.0032	66
Pendimethalin	40487-42-1	Herbicide	Cell division and cell elongation inhibitor	0.0054	90.4
Azoxystrobin	131860-33-8	Fungicide	Mitochondrial respiration inhibitor	0.049	13
Dimethoate	60-51-5	Insecticide, acaricide	Cholinesterase inhibitor	0.14	560

(Pest. Man. 1996). The EC$_{50}$ values are mainly extracted from the US EPA database ECOTOX (2003), but the ECB database IUCLID (2000) and data from RIVM (1999) are also included. Only results from laboratory tests executed under certain standard conditions with freshwater pelagic 'standard' organisms, as described in Appendix B, see OnlineEdition, are included. The comparison of the different average approaches below is based on Larsen et al. (2004), where further details can be found.

5.1 Test for statistical distribution

It is here assumed that the species sensitivity distributions for the substances introduced in Table 1 (overall, within each of the taxa algae, Crustacea and fish, as well as be-

tween the taxa), are lognormal, which is a condition for the use of geometric means. Actual test for goodness of fit by use of the Anderson-Darling test, as described in Aldenberg et al. (2002), requires eight or more data points (D'Agostino1986, quoted by Aldenberg et al. 2002), and is thus only possible at the species and genus level by pooling all the taxa together. The result of the test of goodness of fit is shown in **Table 2**.

Due to lack of data on a sufficient number of species or genera it has only been possible to test six of the substances for lognormal distribution of the species' sensitivity, and only on the total species and total genus level - not within each of the taxa/trophic levels, i.e. algae, Crustacea and fish. As shown in Table 2, lognormal distribu-

Table 2: The number of data values (EC$_{50}$s (acute)) at different levels for each of the included substances. Cases where a goodness of fit test can be performed (i.e. where the number of species or genera ≥ 8) are shown in bold, and if lognormal distribution cannot be rejected at the 5% significance level, the figure is marked with an asterisk (*). The Anderson-Darling test for goodness of fit (modified A^2 test statistics) is used according to Aldenberg et al. (2002, pp 57 and 91) and Stephens (1986, pp 122–125)

Substance	Total number of data [a]	Total number [b]		Number of algae [b]		Number of crustaceans [b]		Number of fish [b]	
		Species	Genera	Species	Genera	Species	Genera	Species	Genera
2,3,4,6-Tetrachlorophenol	12	**9 ***	**8 ***	2	2	2	1	5	5
4-Methyl-2-pentanone	8	4	4	1	1	1	1	2	2
2,4-Dichlorophenol	20	**11 ***	**10 ***	3	2	1	1	7	7
2-Chloroaniline	15	6	5	3	2	1	1	2	2
4-Nitrophenol	35	**10 ***	**9 ***	3	2	1	1	6	6
Dicamba	10	5	5	2	2	1	1	2	2
Metribuzin	18	**11**	**11**	6	6	2	2	3	3
Terbutylazine	11	**10 ***	**10 ***	4	4	1	1	5	5
Pendimethalin	17	6	6	2	2	1	1	3	3
Azoxystrobin	6	6	6	3	3	1	1	2	2
Dimethoate	36	**9 ***	**9 ***	2	2	1	1	6	6

[a] Meaning total number of EC$_{50}$ and LC$_{50}$ data (including data from 'replicates' of tests on the same species)
[b] Meaning total number of different genera or species with EC$_{50}$ or LC$_{50}$ data

tion is only rejected for metribuzin at the 0.05 significance level (α = 0.05). One reason for metribuzin failing the fitness test is a high representation of algae species (six) (which are highly sensitive, EC_{50} = 0.01–0.15 mg/l) as compared to only two species of crustaceans (EC_{50} = 12–35 mg/l) and three species of fish (EC_{50} = 3.4–97 mg/l) for this substance. If the value for the most sensitive species of alga is excluded (i.e. EC_{50} = 0.0117, new, most sensitive EC_{50} = 0.0119), lognormal distribution can no longer be rejected at the 0.05 significance level.

For estimation of the mean of the sensitivity distribution, it is a requirement that the data are normally distributed. Testing the same six pair of data sets for normality reveals that, for all but one (2,4-dichlorophenol), normal distribution is rejected at the 0.05 significance level. For 2,4-dichlorophenol, the sensitivity distribution for species and genera are only just significant at the 0.05 significance level (A^2 (modif.) equals 0.711 and 0.701, respectively, with a threshold limit of 0.752 at the 0.05 significance level) as opposed to the test for lognormality on the same substance, which shows significance at the 0.15 and 0.25 significance level for species and genera, respectively. With the rejection of the normal distribution hypothesis, the use of mean estimates for these tested pairs of data sets is thus, in principle, meaningless for all but one, and the results confirm the observation made in Section 1 that toxicity data are typically lognormally rather than normally distributed.

5.2 Results of the test on real data

As an example, all the data collected for one of the substances, 2,3,4,6-tetrachlorophenol, are shown in **Table 3**. Data, calculations, references, etc. for all 11 substances can be found in Larsen et al. (2004).

As shown in Table 3, the different scenarios (S1, S2, ….) create each their data set by combining either the highest or the lowest measured value from each of the taxa into new data sets each with three values, one for each trophic level. Each scenario could be the outcome of a more limited data availability for this rather well investigated substance, and the idea is to see how robust the different approaches to estimate $HC50_{EC50}$ are, when we – as in most cases – only know three data and not twelve as here. By taking the extreme values within each taxon we obtain the maximum variation within our fictive data sets.

The distributions of the geometric mean (GM), the arithmetic mean (Mean), the median (Median) and the lowest EC_{50} (LowEC$_{50}$) for the eight fictive scenarios in Table 3 are shown in **Fig. 5**. Furthermore, the geometric mean and the median based on either the full data set for all species (sp.) or the geometric means within each taxon (trophic level, troph) are also shown.

From Table 3 and Fig. 5, it can be seen that the geometric mean based on the geometric means for each of the three taxa (i.e. the average on trophic level, termed GM-troph,

Table 3: Data for 2,3,4,6-tetrachlorophenol with geometric means (GM) on level of species, genus and trophic level (mg/l). Eight fictive data availability scenarios (S1-S8) have been constructed with all the possible combinations of highest and lowest measured EC_{50} value from each of the three trophic levels represented by the three taxonomic groups of algae, Crustacea and fish. The mean ecotoxicity effect indicator ($HC50_{EC50}$) has been calculated using the three different average approaches (geometric mean, median and mean). For comparison, the lowest EC_{50}, which would be the basis of a PNEC, has also been determined

Taxon	Species	EC_{50} (acute)	GM-species level	GM-genus level	GM-trophic level	S1	S2	S3	S4	S5	S6	S7	S8
Algae	*Chlorella vulgaris*	10.1	10.1	10.1	3.624	10.1	10.1	10.1	10.1				
	Selenastrum capricornutum	1.3	1.3	1.3						1.3	1.3	1.3	1.3
Crustacea	*Daphnia galeata mendotae*	0.58	0.58	0.312	0.312	0.58	0.58			0.58	0.58		
	Daphnia magna	0.29	0.167										
	Daphnia magna	0.18											
	Daphnia magna	0.09						0.09	0.09			0.09	0.09
Fish	*Poecilia reticulata*	1.085	0.807	0.807	0.475	1.085		1.085		1.085		1.085	
	Poecilia reticulata	0.6											
	Pimephales promelas	1.03	1.03	1.03									
	Oryzias latipes	0.62	0.62	0.62									
	Oncorhynchus mykiss	0.334	0.334	0.334									
	Lepomis macrochirus	0.14	0.14	0.14			0.14		0.14		0.14		0.14
Geometric mean		0.546	0.679	0.749	0.812	1.852	0.936	0.995	0.503	0.935	0.473	0.503	0.254
Median		0.590	0.620	0.713	0.475	1.085	0.580	1.085	0.140	1.085	0.580	1.085	0.140
Mean		1.362	1.675	1.830	1.470	3.922	3.607	3.758	3.443	0.988	0.673	0.825	0.510
LowEC$_{50}$		0.090	0.140	0.140	0.312	0.580	0.140	0.090	0.090	0.580	0.140	0.090	0.090

Fig. 5: Average estimates for the eight fictive scenarios for 2,3,4,6-tetrachlorophenol
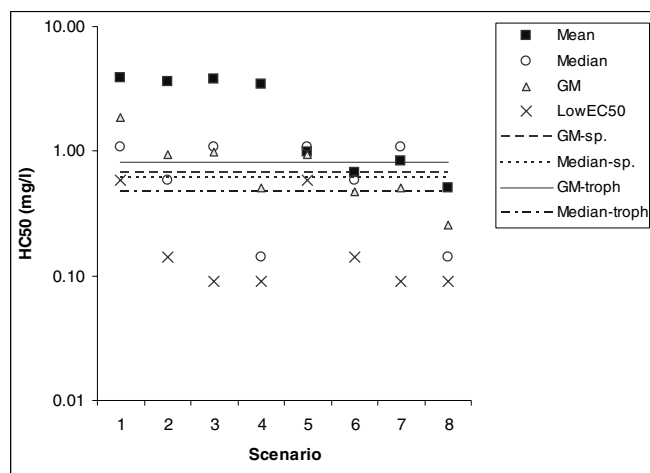

Fig. 6: Geometric means (GMs) with 95% confidence limits for the eight fictive scenarios for 2,3,4,6-tetrachlorophenol

value 0.812) and the geometric mean based on the average of species level (termed GM-sp., value 0.679) are quite close in this case (difference by about a factor of 1.2). The geometric mean based on the average on a genus level (termed GM-genus, value 0.749) is even closer to the GM-troph. The two medians, i.e. the average on the species level (termed Median-sp., value 0.620) and the average on the throphic level (termed Median-troph, value 0.475), are both lower than the GM-troph and GM-sp. (max. difference about a factor of 1.7). But all these different measures of the average based on the full data set are within the 95% confidence limits of the GM-troph (0.0312–21.1) as well as the 95% confidence limits of the GM-sp. (0.256–1.80). These and the following confidence limits are estimated on the basis of the t-statistics according to Campbell (1974, pp 142–144 and Table A12).

As would be expected, the mean values for the eight scenarios are generally higher than the other average measures, only exceeded by the median in two cases (scenario 5 and scenario 7). The GMs for the different scenarios are distributed around the GM-troph and the other averages based on the full data set. This is also the case for the scenario-based medians, but with a small tendency to more extremes (scenario 4 and scenario 8). The scenario values for the LowEC$_{50}$ are the lowest values, as would be expected, but they are actually higher than the Median-troph in two cases (scenario 1 and scenario 5).

If we calculate the ratio between the highest HC50$_{EC50}$ and the lowest HC50$_{EC50}$ estimated within the scenarios for each of the different approaches, we get 7.3 (GM), 7.8 (median), 6.4 (LowEC$_{50}$) and 7.7 (mean). In this respect, they are therefore quite similar.

If we calculate the 95% confidence limits for the GMs, we get the result shown in **Fig. 6**. Here, the confidence limits are quite wide for some of the scenarios (especially scenario 2, 3 and 4) due to a high standard deviation on the GM (large difference between the two extreme values), combined with the fact that the number of data for each is only three.
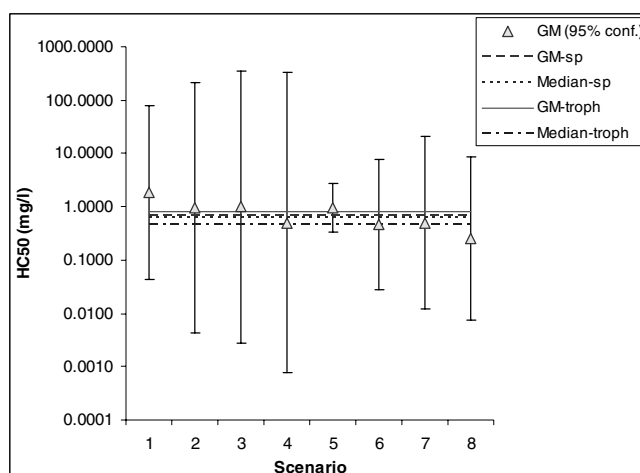
As an alternative, the 90% confidence limits and the 80% confidence limits for the GM may be estimated. The 80% confidence limits for the GM are shown in **Fig. 7**, together with the 95% and the 90% confidence limits for the GM-troph. Even after a reduction of the confidence limits to 80% for this substance, the confidence limits for all the scenarios are still seen to enclose the GM-troph.

Another way to define the limits around the geometric mean is to use the lowest value in the data set (with three observations) as the lower limit and the highest value in the data set as the upper limit. This approach is shown in **Fig. 8**. In this case, the limits for all eight scenarios also enclose the GM-troph, and the width of the limits is almost identical to, or a bit more narrow than, the 80% confidence limits shown in Fig. 7. Also shown in Fig. 8 are the max-min limits around the GM-troph defined as the highest value (upper limit) and the lowest value (lower limit) among the three GM-troph-level values (algae, crustaceans, fish), see Table 3.

The analyses shown here for 2,3,4,6-tetrachlorophenol have been done for all 11 substances shown in Table 1 and re-
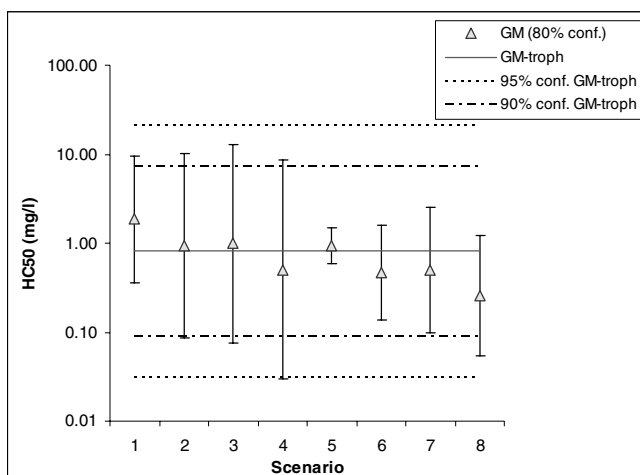

Fig. 7: Geometric means (GMs) with 80% confidence limits for the eight fictive scenarios for 2,3,4,6-tetrachlorophenol
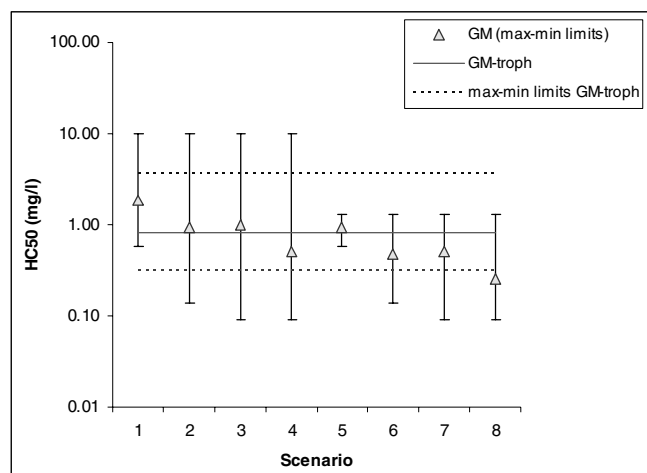
Fig. 8: Geometric means (GMs) with max-min limits for the eight fictive scenarios for 2,3,4,6-tetrachlorophenol

ported in Larsen et al. (2004). For each substance, the ratio between the highest $HC50_{EC50}$ and the lowest $HC50_{EC50}$ estimated within the scenarios for the GM and median approaches has been calculated and shown in **Table 4**. Further, ratios of estimations of $HC50_{EC50}$ on the trophic level and the species level based on the full data set are also shown for both the geometric mean and the median.

The ratios in Table 4 show that the variation in GM is higher than the variation in the median for six of the analysed substances (up to a factor 3.8). For the remaining five substances, the variation in the median is highest. In most cases, the ratio of the variation in the median to the variation in the GM lies below a factor 2, but, in one case (dimethoate), the ratio is close to 10 (840/85). The example of dimethoate illustrates the sensitivity of the median when a shift occurs in a three value data set from two low values and one high, to two high values and one low as shown theoretically in Fig. 2 for a five value data set ('jumping median'). This strong effect is potentially a problem for substances with a highly specific TMoA.

The calculated averages based on the full data set (geometric mean or median) on the species level and the trophic level are in most of the cases quite close in Table 4, but, in some of the investigated cases, the difference in the number of species represented at each trophic level leads to a large difference in average toxicity. One example is metribuzin for which the GM-troph is nearly four times higher than the GM-sp. (or GM-genus, not shown). The main reason for this difference is that the substance is especially toxic to algae which are represented by six species/genera, whereas crustaceans are only represented by two species/genera, and fish only by three species/genera. This effect becomes even more pronounced when comparing the Median-sp. or Median-genus (not shown) with the Median-troph, thus giving rise to a ratio of 140 in difference. Further examples and discussions of differences between GM-sp. and GM-troph can be found in Larsen (2004).

Examples of substances with a behaviour similar to the theoretical case shown in Fig. 1 are provided by 4-nitrophenol and metribuzin, although the medians are not identical for these substances. On a three value data set representing geometric means at each trophic level (14.6; 12.0; 11.6 for 4-nitrophenol and 30.1; 20.6; 0.038 for metribuzin), the median approach gives the values 12.0 (4-nitrophenol) and 20.6 (metribuzin), indicating that 4-nitrophenol on average is more toxic than metribuzin despite the fact that metribuzin is very toxic to algae. The geometric mean, on the other hand, gives the values 12.7 (4-nitrophenol) and 2.86 (metribuzin) indicating a higher average toxicity of metribuzin.

Summarizing the different approaches to calculate (confidence) limits around the geometric mean for all 11 substances, it can generally be concluded that:

- 95% confidence limits around a GM-troph are in most cases very wide making the differentiation between the average toxicity among chemicals impossible
- Though more narrow than the 95% confidence limits, the 90% confidence limits around GM-troph are still relatively wide

Table 4: Ratio between highest and lowest scenario-estimated $HC50_{EC50}$ and full data set-based ratios for substances included in the test of different 'average approaches'

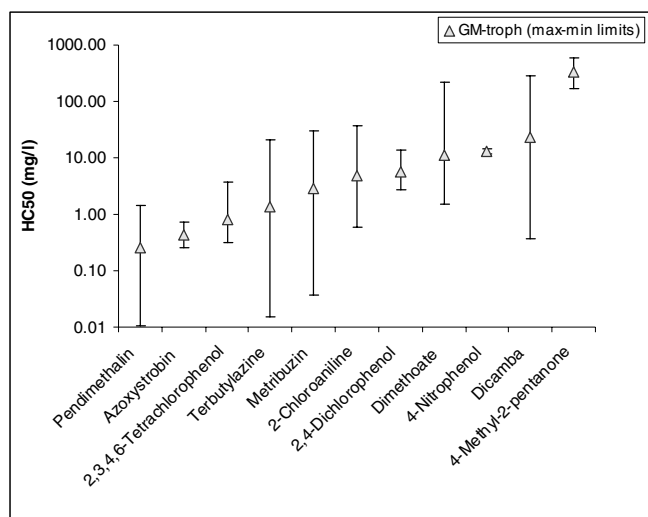| Substance name	Scenario-based ratios (highest HC50/lowest HC50)		Full data set-based ratios (HC50/HC50)	
	GM	**Median**	**GM-troph/GM-sp.**	**Median-troph/Median-sp.**
2,3,4,6-Tetrachlorophenol	7.3	7.8	1.2	0.77
4-Methyl-2-pentanone	1.2	1.0	0.87	0.83
2,4-Dichlorophenol	3.8	6.6	1.1	0.76
2-Chloroaniline	4.2	1.1	0.49	0.29
4-Nitrophenol	10	7.7	0.86	0.94
Dicamba	12	6.4	1.6	1.3
Metribuzin	24	29	3.9	140
Terbutylazine	11	13	1.9	2.7
Pendimethalin	38	37	0.97	1.2
Azoxystrobin	8.5	4.2	0.93	1.1
Dimethoate	85	840	1.2	0.62

Fig. 9: GM-troph with max-min limits for the 11 included substances (mg/l)

- Even the 80% confidence limits around the GM-troph are in most cases a bit wider than the limits based on the min-max values of the three data GM-troph data set (based on one value from each trophic level)
- The limits based on min-max values are relatively narrow for the substances tested here and, in all but one case (i.e. 4-nitrophenol, not shown), all eight fictive scenarios overlap with the GM-troph calculated on the basis of the full data set (see example in Fig. 8)
- In the case of 4-nitrophenol (not shown) not even the 95% confidence limits around the GM (scenario 1 and scenario 8) overlap with the GM-troph and, for these two extreme scenarios, the estimated GMs are 45 mg/l and 4.4 mg/l and with a GM-troph value of 13 mg/l, which gives rise to an 'error' of a factor of 3-4.

The GM-troph and the associated min-max limits for all 11 substances are shown in **Fig. 9.**

6 Conclusions, Recommendations and Perspectives

On the basis of theoretical and empirical analyses of the different average approaches, including different ways of calculating (confidence) limits for the data sets of three observations (one from each of the three trophic levels), the following recommendations are given for the calculation of EEIs to be used in life cycle impact assessment:

- The indicator should be based on the GM-troph calculated as the geometric mean of the three EC_{50} values, one from each trophic levels represented by algae, invertebrates (crustaceans) and fish.
- If more than one EC_{50} value from each trophic level is available, the GM-troph should then be calculated as the geometric mean of the geometric means for each trophic level (GM-trophic-levels). The GM-trophic-levels are calculated as the geometric mean of the GM-genus-level, which again are calculated as the geometric mean of the GM-species-level which, in their turn, are calculated as the geometric mean of the single EC_{50} values for each species, as illustrated in the example for 2,3,4,6-tetrachlorophenol in Table 3.

- As limits around the GM-troph, the lowest EC_{50} value should be used as the lower limit, and the highest EC_{50} value as the upper limit in data sets with only three EC_{50} values, i.e. one from each trophic level.
- If more than one EC_{50} value from each trophic level is available, the max-min limits around the GM-troph should be based on the three GM-trophic-level values, i.e. the lowest GM-trophic-level value is used as the lower limit and the highest GM-trophic-level value as the upper limit.

By using the GM-troph it is suggested to put equal weight on each of the three trophic levels (represented by algae, crustacean and fish) instead of relying on an arbitrary species representation where the weights put on each trophic level indirectly is determined by regulatory priorities of a dubious ecological relevance. By making a conscious choice of equal weights to each of the trophic levels, we try to avoid the possible bias from data sets with a highly unequal number of tested species/genera among the three trophic levels, which would occur if GM-sp. (or GM-genus) were used instead.

In order to secure a broader coverage of biological sensitivity, inclusion of other taxa than algae, crustaceans and fish like insects, mollusca and amphibians may be relevant. However, no standard tests (freshwater, pelagic) currently exist for these taxa (Larsen et al. 2004), and the availability of test results is highly variable among chemicals. So, what has been done here is a trade off between on the one side seeking a broad coverage of biological sensitivities with the risk of introducing a bias in the weighting of chemical ecotoxicity effect indicators as compared to one another and, on the other side, seeking a robust standardised basis for comparison, although then running the risk of excluding very sensitive species. Taking into account the comparative context of LCA, seeking best estimate and equal treatment of a high number of chemicals, and further considering the fact that non-standardised test data are most variable in quality and availability among chemicals, as compared to acute tests on standard organisms, the choice of exclusively using results from tests on standard organisms, following certain restrictions on end point and time duration, seems most reasonable. However, the importance of excluding very sensitive species should be investigated and is therefore part of the recommended further research on ecotoxicity effect indicators.

As suggested by Forbes and Calow (2002a), a way to increase the ecological relevance, when dealing with the species sensitivity distribution approach, might be to assign weights to the input values from each taxon (or trophic level) which reflect the relative abundance of different taxa in the community/ecosystem in question. Combining this approach with theories of functional redundancy of species (Pratt and Cairns 1996) could be a very interesting research area for the further development of the EEI.

It is not recommended to determine a statistically estimated uncertainty as a basis for the limits around the GM-troph, even though the number of input data is higher than three in some cases. This is in order to use the same approach for all substances and because the max-min approach is simple and seems to work well, at least for the 11 rather diverse sub-

stances tested here. Furthermore, in most cases, only three relevant data values are available anyway. The results obtained here indicate that the max-min range of the three-data GM-troph (no matter the database) in practice includes the 'true' GM-troph (based on the full data set), as well as the confidence limits. Even though the max-min limits cannot be used for testing statistically significant differences between EEIs, they may be used for giving reasonable certainty that the 'true' GM-troph is included.

The test of the different average approaches undertaken here is based on 11 different substances comprising seven different TMoA. To substantiate the general validity of the conclusions, a higher number of substances covering a larger number of TMoA should be tested.

It is recommended to use EC_{50}(chronic) values when possible, but, as only acute data will be available in most cases, the use of best estimate assessment factors is recommended to extrapolate from acute to chronic values. Even though there is a need for research in this area, an acute to chronic ratio of 2 between $HC50_{EC50}$(acute) and $HC50_{EC50}$(chronic) is recommended as a starting point.

Because of the comparative framework of LCIA seeking best estimates, it is recommended only to use test results from laboratory tests, fulfilling certain standard conditions, e.g. standard organism and restrictions on test duration and endpoints as described in Appendix B (see OnlineEdition), when estimating the $HC50_{EC50}$ value. These standard conditions are described here for acute tests, but detailed criteria for choice of chronic data still need to be developed.

The ability of a geometric mean to represent the toxicity (including chronic toxicity) of very toxic substances and very sensitive species has not been dealt with yet, and further research is needed. However, it may be anticipated, on the basis of the results from the tests performed here of different average approaches on 11 substances (including very toxic pesticides, e.g. terbutylazine), that the GM-troph with its max-min limits to some degree also accounts for very toxic substances if representative toxicity data are available.

In summary, tests on a wider range of substances should be performed to further verify the GM-troph approach and its ability to represent substances that are very toxic to specific species. Also, further research on detailed criteria for choice of chronic test data should be included as well as studies on best estimate assessment factors for acute to chronic extrapolation.

References

Aldenberg T, Jaworska J, Traas TP (2002): Normal Species Sensitivity Distributions and Probabilistic Ecological Risk Assessment. Chapter 5. In: Posthuma L, Suter II GW, Traas TP (eds), Species Sensitivity Distribution in Ecotoxicology. Lewis, Boca Raton, FL, USA, pp 49–102

Bliss CI (1935): The calculation of the dosage-mortality curve. Ann Appl Biol 22, 134–167

Cairns JJr, Niederlehner BR, Smith EP (1995): Ecosystem Effects: Functional End Points. Chapter 19. In: Rand GM (ed), Fundamentals of Aquatic Toxicology. Taylor & Francis, Philadelphia, USA, pp 589–607

Calow P (1998): General Principles and Overview. Chapter 1. In: Calow P (ed), Handbook of Ecotoxicology. Blackwell Science, Oxford, Great Britain, pp 1–5

Campbell RC (1974): Statistics for Biologists. Second edition. Cambridge University Press, Cambridge, Great Britain

Chapman PF, Crane M, Wiles JA, Noppert F, McIndoe EC (1996): Asking the Right Questions: Ecotoxicology and Statistics. SETAC, Belgium. The report of a workshop held in Royal Holloway University of London, Egham, Surrey, United Kingdom, 26–27 April 1995

Chapman PM, Wang F, Janssen CR, Goulet RR, Kamunde CN (2003): Conducting ecological risk assessments of inorganic metals and metalloids: Current status. Human Ecol Risk Assess 9, 641–697

D'Agostino RB (1986): Tests for normal distribution. In: D'Agostino, RB, Stephens MA, (eds), Goodness-of-Fit Techniques. Marcel Dekker, New York, USA, pp 367–419

EC (2003a): Environmental Risk Assessment. Chapter 3 (Part II) In: Technical Guidance Document on Risk Assessment in support of Commission Directive 93/67/EEC on Risk Assessment for New Notified Substances, Commission Regulation (EC) No. 1488/94 on Risk Assessment for Existing Substances and Directive 98/8/EC of the European Parliament and the Council concerning the placing of biocidal products on the market. Institute for Health and Consumer Protection. European Commission. ECB, JRC, Ispra, Italy

EC (2003b): Proposal for a Regulation of the European Parliament and of the Council concerning the Registration, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency and amending Directive 1999/45/EC and Regulation (EC) {on Persistent Organic Pollutants}. COM (2003) 644, Brussels, 29 October 2003

EEC (1967): Approximation of laws, regulations and administrative provision relating to the classification, packaging and labelling of dangerous substances (Council Directive 1967/548/EEC, June 27)

ECOTOX (2003): ECOTOXicology Database system. US EPA, <http://www.epa.gov/ecotox/>

Forbes VE, Calow P (2002a): Species Sensitivity Distribution Revisited: A Critical Appraisal. Human and Ecological Risk Assessment 8 (3) 473–492

Forbes VE, Calow P (2002b): Extrapolating in Ecological Risk Assessment: Balancing Pragmatism and Precaution in Chemical Controls Legislation. BioScience 52 (3) 249–257

Goedkoop M, Spriensma R (2001a): The Eco-indicator 99. A damage oriented method for Life Cycle Impact Assessment. Methodology Report. Third edition. Pré Consultants, Amersfoort, The Netherlands <http://www.pre.nl>

Goedkoop M, Spriensma R (2001b): The Eco-indicator 99. A damage oriented method for Life Cycle Impact Assessment. Methodology Annex. Third edition. Pré Consultants, Amersfoort, The Netherlands <http://www.pre.nl>

Hauschild M (2005): Assessing environmental impacts in a life cycle perspective. Environmental Science & Technology 39 (4) 81A–88A

Hauschild M, Pennington D (2002): Indicators for Ecotoxicity in Life-Cycle Impact Assessment. Chapter 6. In: Udo de Haes HA, Finnveden G, Goedkoop M, Hauschild M, Hertwich E, Hofstetter P, Jolliet O, Klöpffer W, Krewitt W, Lindeijer E, Müller-Wenk R, Olsen SI, Pennington D, Potting J, Steen B (eds), Life-cycle Impact Assessment: Striving Towards Best Practice. SETAC-Europe, Brussels, Belgium, pp 149–176

Hauschild M, Wenzel H, Damborg A, Tørsløv J (1998): Ecotoxicity as a criterion in the environmental assessment of products. Chapter 6. In: Hauschild M, Wenzel H (eds), Environmental Assessment of Products: Volume 2: Scientific background. Chapman & Hall, London, Great Britain, pp 203–314

Huijbregts M, Thissen U, Guinée JB, Jager T, Kalf D, Van de Meent D, Ragas AM J, Wegener Sleeswijk A, Reijnders L (2000): Priority assessment of toxic substances in life cycle assessment. Part I: Calculation of toxicity potentials for 181 substances with the nested multi-media fate, exposure and effects model USES-LCA. Chemosphere 41, 541–573

ISO (2000): ISO 14042: Environmental management – Life cycle assessment – Life cycle impact assessment. International Organisation for Standardisation, Geneva, Switzerland

Kusk O (2003): Personal communication with Ole Kusk, Environment and Resources, Technical University of Denmark

IUCLID (2000): European Chemical Bureau: IUCLID CD-ROM Year 2000 edition. Public data on high volume chemicals. ISBN 92-828-8641-7

Jolliet O, Margni M, Charles R, Humbert S, Payet J, Rebitzer G, Rosenbaum R (2003): Impact 2002+: A New Life Cycle Impact Assessment Methodology. Int J LCA 8 (6) 324–330

Larsen HF, Hauschild M (2006): Evaluation of ecotoxicity effect indicators for use in LCIA. Int J LCA DOI: <http://dx.doi.org/10.1065/lca2006.12.287>

Larsen HF, Hansen MS, Hauschild M (2006): Life-cycle assessment of offset printed matter – How important are the chemical-related impacts? Submitted to Journal of Cleaner Production

Larsen HF (2004): Assessment of chemical emissions in life cycle impact assessment – Focus on low substance data availability and ecotoxicity effect indicators. Ph.D. Thesis, October 2004. Department of Manufacturing, Engineering and Management. Technical University of Denmark, ISBN 87-91035-22-8. <http://www.tempo.ipl.dtu.dk/ipl/upload/publ/PhD-thesis-rev.pdf>

Larsen HF, Payet J, Molander S, Hauschild M (2004): Implementation of the OMNIITOX Base Model. Contribution to work package 8 of the OMNIITOX project as part A of appropriate deliverable D41. Part VIII – Implementation of the ecotoxicological effects module. EC Project contract G1RD-CT-2001-00501 <http://omniitox.imi.chalmers.se/OfficialMirror>

Lewis MA (1998): Freshwater Primary Producers. Chapter 3 In: Calow P (ed), Handbook of Ecotoxicology. Blackwell Science, Oxford, Great Britain, pp 28–45

Mayfield CI (1998): Microbial Systems. Chapter 2 In: Calow P (ed), Handbook of Ecotoxicology. Blackwell Science, Oxford, Great Britain, pp 9–27

Newman MC, Dixon PM (1996): Ecologically Meaningful Estimates of Lethal Effect in Individuals. Chapter 8. In: Newman MC, Jagoe CH (eds), Ecotoxicology – A Hierarchical Treatment. Lewis, Boca Raton, FL, USA, pp 225–253

Newman MC, Ownby DR, Mézin LCA, Powell DC, Christensen TRL, Lerberg SB, Anderson B-A, Padma TV (2002): Species Sensitivity Distributions in Ecological Risk Assessment: Distributional Assumptions, Alternate Bootstrap Techniques, and Estimation of Adequate Number of Species. Chapter 7. In: Posthuma L, Suter II GW, Traas TP (eds), Species Sensitivity Distributions in Ecotoxicology. Lewis, Boca Raton, FL, USA, pp 119–132

OECD (1998): Detailed Review Paper on Aquatic Testing Methods for Pesticides and Industrial Chemicals. Part 1: Report and Part 2: Annexes. Series on Testing and Assessment No. 11. OECD, Paris, France

OECD (2003): Manual for investigations of HPV chemicals. Chapter 2: SIDS, the SIDS plan and the SIDSA dossier. OECD Existing Chemicals Programme <http://www.oecd.org>

Payet J, Larsen HF (2006): Endpoint modelling for the life cycle impact assessment on ecosystems. (in preparation)

Payet J (2005): Assessing Toxic Impacts on Aquatic Ecosystems in LCA. Int J LCA 10 (5) 373

Payet J, Jolliet O (2005): Comparative Assessment of the Toxic Impact of Metals on Aquatic Ecosystems: The AMI method. In: Dubreuil A (ed), Life-Cycle Assessment of Metals: Issues and research directions. (Proceedings of the International Workshop on Life-Cycle Assessment and Metals. Montreal, Canada. 15–17 April 2002). SETAC, Pensacola, FL, USA

Payet J (2004): Assessing Toxic Impacts on Aquatic Ecosystems in Life Cycle Assessment (LCA). PhD thesis No. 3112 (2004). EPFL, Lausanne, Switzerland

Payet J, Pennington D, Hauschild M, Jolliet O (2002): The AMI method for the assessment of the median impact on ecosystems: Description and guidelines. EPFL, Lausanne, Switzerland

Pennington D, Payet J, Hauschild M (2004): Aquatic ecotoxicological indicators in life-cycle assessment. Environ Toxicol Chem 23 (7) 1796–1807

Pest. Man. (1996): The Pesticide Manual. 11. ed. Book. ISBN 1 901396 11 8. The British Crop Protection Council, 49 Downing Street Farnham, Surrey GU9 7PH, UK

Pratt JR, Cairns JJr (1996): Ecotoxicology and the Redundancy Problem: Understanding Effects on Community Structure and Function. Chapter 12 In: Newman MC, Jagoe CH (eds), Ecotoxicology – A Hierarchical Treatment. Lewis, Boca Raton, FL, USA, pp 347–370

RIVM (1999): Appendix A. In: Könemann H (ed), Environmental Risk Limits in The Netherlands. Part III Data. Report No. 601640 001, RIVM, Bilthoven, The Netherlands

Russom CL, Bradbury SP, Broderius SJ, Hammermeister DE, Drummond RA (1997): Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow (*Pimephales promelas*). Environ Toxicol Chem 16 (5) 948–967

Sand-Jensen K (2000): Økologi og biodiversitet. Overordnede mønstre for individer, bestande og økosystemer. Gads forlag, København, Danmark

Stephens MA (1986): Tests based on EDF statistics. In: D'Agostino RB, Stephens MA (eds), Goodness-of-Fit Techniques. Marcel Dekker, New York, USA, pp 97–193

Solbé JF, De LG (1998): Freshwater Fish. Chapter 5. In: Calow P (ed), Handbook of Ecotoxicology. Blackwell Science, Oxford, Great Britain, pp 66–82

Udo de Haes HA, Lindeijer E (2002): The conceptual Structure of Life-Cycle Impact Assessment. Chapter 8. In: Udo de Haes HA, Finnveden G, Goedkoop M, Hauschild M, Hertwich E, Hofstetter P, Jolliet O, Klöpffer W, Krewitt W, Lindeijer E, Müller-Wenk R, Olsen SI, Pennington D, Potting J, Steen B (eds), Life-cycle Impact Assessment: Striving Towards Best Practice. SETAC-Europe, Brussels, pp 209–225

US EPA (2000): Data collection and development on high production volume (HPV) chemicals. Federal Register 65 (248) 81686–81698 <http://www.epa.gov/chemrtk/>

Versteeg DJ, Belanger SE, Carr GJ (1999): Understanding single-species and model ecosystem sensitivity. Data-based comparison. Environ Toxicol Chem 18, 1329–1346

Whittaker RH (1975): Communities and Ecosystems. Second edition. Macmillan, New York, USA

Appendix A. Environmental Compartments

For an EEI, a distinction is typically made between at least the two compartments water and soil, as in the EDIP method (Hauschild et al. 1998) and the Eco-indicator 99 method (Goedkoop and Spriensma 2001a, 2001b). In other LCIA methods like USES-LCA (Huijbregts 2000), a more detailed distinction is made. Water is thus divided into freshwater and saltwater, which is further divided into a pelagic and a sediment part, and soil may be distinguished according to its use into e.g. agricultural soil, industrial soil and natural soil. In risk assessment, a distinction is typically also made between the aquatic compartment (pelagic) and the sediment (benthic) as described in the TGD (EC 2003a). From an ecological point of view, the pelagic and the benthic (sediment) compartment are quite different with different types of organisms, but in most cases with high interaction.

In order not to create bias in the effect indicator by mixing up two quite different habitats/compartments (e.g. due to lack of sediment data in many cases), it is recommended to distinguish between the pelagic and the sediment compartment. Tests on benthic or sediment dwelling organisms should be included in a separate sediment compartment. The same basic estimation principle as for the freshwater pelagic compartment may be used, and lack of toxicity data could as a starting point be solved by use of the equilibrium partitioning method (EC 2003a). For the marine compartment, at least separated into a pelagic and a sediment compartment, and for the soil (terrestrial) compartment, the same approach may be used, but based on saltwater organisms and soil organisms, respectively. However, the inclusion of the marine and soil compartment needs further research if the average PAF approach (i.e. GM-troph) is going to be applied, and this lies outside the scope of this paper.

Appendix B. Ecotoxicity Tests

Appendix B.1. Overall performance criteria

Calow (1998) defines five performance criteria for ecotoxicity tests:

- Relevance, ecological realism
- Reproducibility
- Reliability
- Robustness
- Repeatability/sensitivity

In LCIA, we are integrating impacts over time and space, so, in that sense, we are far from the ecological realism of the potential effect of the integrated emissions we estimate. This restriction, dictated by the nature of the life cycle or product system, and the fact that the LCA is focused on a functional unit rather than full output from the processes (Hauschild 2005), points in the direction of not putting highest priority on ecological relevance. The fact that EEI is to be used in the comparative framework of LCIA, i.e. for comparison of emitted substances, puts the focus on reproducibility, robustness and repeatability/sensitivity.

What we are dealing with in LCIA is anticipatory tests (as opposed to assessment tests), where reproducibility is important (Calow 1998). On the other hand, if we are dealing with site specific LCIA, we will approach the relevance of using assessment tests because ecological realism may become more important.

As our goal is to treat all chemicals equally, we should aim at choosing a set of tests reflecting the sensitivity of representative 'standard' organisms from at least three trophic levels. This should be done in a way where the knowledge of today is taken into account, trying to avoid bias due to differences in the sensitivity of haphazard test organisms included. For most of the chemicals to be modelled by the EEI, we only have access to three acute data values, i.e. $E(L)C_{50}$ values for fish, Daphnia and algae, as defined, for example, by the proposal for a new EU legislation on chemicals, i.e. REACH, EC (2003b) and the OECD work on investigations of high production volume chemicals (OECD 2003). Inclusion of non-standardised test result values in the calculation of the effect indicators, in those cases where it is possible, will increase variance and most probably create bias because the type and availability of such data is very variable among chemicals.

Furthermore, as illustrated by Newman and Dixon (1996), when comparing the predominant dose or concentration response methods (time endpoint methods) with time response methods (time to death, TTD methods), exposure time and covariance from factors like sex, bodyweight and genotype can have significant influence on the acute lethal EC_{50} determined in a laboratory test. These results strengthen the arguments for only using results from standardised tests when seeking best estimates in a comparative framework based on few available data.

There are thus good reasons to aim at basing the effect indicator exclusively on results from use of standardised test guidelines and standardised test organisms. However, we recognize that this kind of information in many cases is not directly available from the data sources typically used, e.g. ECOTOX (2003), and many test data which could be very useable (if reliable) are not produced under conditions strictly following a standardised test guideline. Therefore, besides the standard test guidelines, focus in the following sections will also be on test species and some other of the key test criteria to be used when choosing relevant data input.

Appendix B.2. Specific demands on ecotoxicity tests

If we look at standard freshwater laboratory tests, single species tests are dominating and the results are to a large extent used for regulatory purposes (e.g. water quality criteria) as reflected in the EU legislation on chemicals, i.e. Annex V of 67/548/EEC (EEC 1967) and in the USA, for example, in the 'Chemical Right-to-Know Initiative' on high production volume chemicals (US EPA 2000). A lot of work on standardisation, evaluation, etc. on single species labo-

ratory tests has therefore been done, and ecotoxicity test data come almost exclusively from this source.

Below, is a short assessment and recommendation of which key test criteria (i.e. endpoint, species and test duration) to focus upon when choosing acute data for primary producers, primary consumers and secondary consumers, to be included in the estimation of the EEI. For further evaluation of acute tests and key test conditions, including recommendations on standard test guidelines, see Larsen et al. (2004).

Appendix B.2.1. Primary producers

Among standard laboratory tests with primary producers, tests with freshwater algae are dominating. Even though vascular plants play an important role in a freshwater ecosystem (e.g. as food and shelter), only a few standardised methods exist (Lewis 1998) and, due to the limited amount of test results, the sensitivity of these species is not as well understood as the sensitivity of algae. So, for freshwater primary producers other than algae, standard tests either do not exist or the number of test results is very limited. It is therefore assessed that inclusion of non-algae test results may create bias in the calculation of the average toxicity of chemicals due to a lack of data for the main part. Therefore, it is recommended that the test results on primary producers to be included in the calculation of the EEI only include algae.

Lewis (1998) compiles existing standard test guidelines for algae. Almost all of these and a few more are assessed in the OECD Detailed Review Paper (DRP) on aquatic testing methods (OECD 1998). On this basis, six standard algae test methods are chosen and compiled in Larsen et al. (2004).

The endpoints used for algae tests include inhibition of growth and photosynthesis. Growth inhibition can either be measured as a reduction in biomass production or as a reduction in growth rate. No scientific consensus exists on which to choose so both biomass and growth rate are acceptable (Lewis 1998). However, according to Kusk (2003), the use of growth rate is dominating today. One of the reasons for this is that a reduction in growth rate is relatively independent of the test time (as long as the growth is exponential) as compared to reduction in biomass which is dependent on the final biomass at the end of the test period (Kusk 2003). Effects on photosynthetic activity seem to be less sensitive to toxic exposure than inhibition of growth in most cases (Lewis 1998). It is therefore recommended to use test results based on inhibition of growth, i.e. either biomass or growth rate, as data input in the calculation of average toxicity to algae.

Algae tests on growth rate are multi-generation tests and they are therefore, in principle, chronic (or long-term) tests. However, proliferation is typically due to mitosis (no sexual reproduction) and EC_{50} values are within generic risk assessment, according to the TGD (EC 2003a), considered as acute values (or short-term values), whereas NOEC values from the same test are used as chronic values or long-term values. For the EEI, we recommend using the EC_{50} as both an acute value and as a chronic value.

The most frequently used algae species in standard laboratory tests are the two green algae *Raphidocelis subcapitata* (former name *Selenastrum capricornutum*) and *Scenedesmus subspicatus*, but also blue-green algae and diatoms are used. The sensitivity of the different algae species may vary more or less depending on the chemical tested. In some cases, more than a factor of 100 or even more than a factor of 1000 is found between species, especially for pesticides and some metals. Unfortunately, the response to chemicals is unpredictable (Lewis 1998). A way to deal with this variation is to include test results on different species, if available. The recommended species include: *Raphidocelis subcapitata*, *Scenedesmus subspicatus*, *Scenedesmus quadricauda*, *Chlorella vulgaris*, *Anabaena flos-aqua*, *Microcystis aeruginosa*, *Navicula seminulum* and *Navicula pelliculosa*.

Typical test duration for a standard algae test is 72–96 hours chosen to obtain exponential growth during the test period. This is also the duration of test methods aiming at estimating EC_{50} for growth inhibition (Lewis 1998). The TGD (EC 2003a) states that it is generally accepted that EC_{50} values from a test duration of 72 hours or longer may be considered as results from a short-term test. Of all standard tests, as compiled in Larsen et al. (2004), only the FIFRA test has a duration above 96 hours, viz. 120 hours. It is therefore recommended that a test duration of 72 hours–120 hours is used for the EEI.

Appendix B.2.2. Primary consumers

Invertebrates play a very important role as part of the primary consumers in the aquatic ecosystem. In standard laboratory tests, the crustaceans (phylum Arthropoda) are dominating, especially the genus *Daphnia* with the species *D. magna* and *D. pulex* accounting for the major part of the test results on invertebrates. As is the case for algae, these kinds of tests are to a large extent used for regulatory purposes. The organisms used for laboratory tests on primary consumers almost exclusively belong to the invertebrates.

As described in the OECD DRP (OECD 1998) on aquatic testing methods on herbivores (and omnivores), i.e. primary consumers, a lot of non-standardised methods making use of a wide range of non-standardised organisms are published. These are all invertebrates and the dominating taxonomic phyla include Arthropoda (Crustacea, Insecta), Protozoans (e.g. *Paramecium*), Rotifera (e.g. *Brachionus*), Cnidarians (*Hydra*), Platyhelminthes (*Dugesia*) and Mollusca. For acute tests, only six species with matching standard test guidelines are given the rating A (or AA) in the OECD DRP (OECD 1998), meaning that these tests are international standards, or international draft standards or national standards that have been subject to national (or international) ring-testing. These species with adjacent test methods are compiled in Larsen et al. (2004). The six species which are recommended for the EEI are: *Daphnia magna*, *Daphnia pulex*, *Daphnia sp.*, *Ceriodaphnia dubia*, *Neomysis mercedis* and *Brachionus calyciflorus*.

The endpoint typically used in acute tests on invertebrates is mortality or immobility (*Daphnia*). Immobility is used in tests on *Daphnia* because it is very difficult to determine when the organism is actually dead. For the EEI, an EC_{50} value based on mortality or immobility is recommended.

For the test duration, it is recommended to use 24–96 hours (48 hours preferred) for the genera *Daphnia* and *Ceriodaphnia*. For *Neomysis*, a test duration of 96 hours is recommended and for *Brachionus* 24 hours.

Appendix B.2.3. Secondary consumers

Fish play a major role as part of the secondary (and tertiary) consumers in the aquatic ecosystem. In standard laboratory tests on carnivores, tests on fish belonging to the superorder Teleostei ('bony fishes') are dominating, especially species like rainbow trout (*Onchorhynchus mykiss*) and bluegill sunfish (*Lepomis macrochirus*) are used. As is the case for algae and invertebrates, these kinds of tests are to a large extent used for regulatory purposes.

According to the OECD DRP on aquatic testing methods (OECD 1998), only tests on carnivorous fish have been published and these tests are dominating the group of non-herbivorous consumers. So, besides tests on fish only a few carnivorous species from each of the phyla Protozoans (e.g. *Para-mecium*), Arthropoda (Insecta), Platyhelminthes (*Dugesia*) and the taxon Amphibia are represented. For acute tests, only 14 species with matching standard test guidelines are given the rating A (or AA) as both an overall score and on standardisation in the OECD DRP (OECD 1998). These are all fish tests and compiled together with adjacent test methods in Larsen et al. (2004). The 14 species, which are recommended for the EEI, are: *Ambassis macleayi, Carassius auratus auratus, Cyprinus carpio carpio, Danio rerio (Brachydanio rerio), Ictalurus punctatus, Lepomis cyanellus, Lepomis macrochirus, Leuciscus idus, Melanotaenia splendida inornata, Onchorhynchus kisutch, Onchorhynchus mykiss (Salmo gairdneri), Oryzias latipes, Pimephales promelas, Poecilia reticulate* and *Salvelnius fontinalis*.

The endpoint typically used in acute tests on fish is mortality. For the EEI, an LC_{50} value is recommended.

For the test durations, it is recommended to use 96–336 hours (96 hours preferred) for the EEI. A test duration of 96 hours (short-term acute test, not including feeding) is used in all recommended standard tests (i.e. given the rating A (or AA), see above) except for two (OECD 1998). In these two tests, which are long-term acute tests (feeding included), the measure of an acute effect (end point LC_{50}) is included and the exposure period is short (i.e. acute) as compared to the life span of the organisms (OECD 1998, Solbé 1998).